4182–4197 *Nucleic Acids Research, 2010, Vol. 38, No. 12* *Published online 1 March 2010*
doi:10.1093/nar/gkq088

Robustness and modularity properties of a non-covalent DNA catalytic reaction

David Yu Zhang* and Erik Winfree

California Institute of Technology, MC 136-93, 1200 E. California Blvd., Pasadena, CA 91125, USA

Received November 18, 2009; Revised January 29, 2010; Accepted February 2, 2010

ABSTRACT

The biophysics of nucleic acid hybridization and strand displacement have been used for the rational design of a number of nanoscale structures and functions. Recently, molecular amplification methods have been developed in the form of non-covalent DNA catalytic reactions, in which single-stranded DNA (ssDNA) molecules catalyze the release of ssDNA product molecules from multi-stranded complexes. Here, we characterize the robustness and specificity of one such strand displacement-based catalytic reaction. We show that the designed reaction is simultaneously sensitive to sequence mutations in the catalyst and robust to a variety of impurities and molecular noise. These properties facilitate the incorporation of strand displacement-based DNA components in synthetic chemical and biological reaction networks.

INTRODUCTION

Nucleic acids have been identified and demonstrated as versatile nanoscale engineering materials because of their specific binding properties and their well-understood thermodynamics (1) and kinetics (2,3). Recent constructions exhibiting dynamic behavior include logic gates and networks (4–6), motors (7,8), and amplification mechanisms (9–15). By perfecting the design of these and other modular primitives, nucleic acid engineering may one day allow precise spatial–temporal control of chemistry (16–19). Such 'molecular programming', if made sufficiently robust, could be incorporated into biological cells and organisms to allow the dynamic programming of development and behavior.

In order for such dreams to be realized, however, the basic nucleic acid components being developed must approach a level of modularity and robustness comparable with those of transistors and other analogs from electrical engineering. Robustness denotes a low amount of interference between the device and the environment. For aqueous molecular computation applications, robust nucleic acid constructions must (i) function over a broad range of ambient solution conditions, (ii) function despite imperfections and impurities and (iii) function in the presence of external molecular noise (in the form of other biomolecules that may exist in solution). Modularity denotes a low amount of interference between devices. For nucleic acid devices, modularity requires a generalized, algorithmic method of component construction such that different components interfere with each other minimally. In this work, we explore and evaluate the robustness and modularity properties of one particular DNA component, so that we can appropriately incoporate it in large and/or noisy systems.

One key component needed for generalized molecular computation is the signal amplifier. In biology as well as in electrical engineering, signal amplification serves a key role in the robustness of networks and kinetics of signal transduction. Nucleic acid signal amplifiers have been implemented on a molecular level with the construction of non-covalent catalytic reactions, in which single-stranded molecules of DNA (henceforth 'strands') catalytically release product strands from multi-stranded complexes (9,13–15). To concisely describe the functions of and reactions involving each strand, we subdivide strands into 'domains', contiguous bases that act as an abstract unit in binding and dissociation (Figure 1A) (9).

Figure 1B shows the designed pathway of the catalytic reaction presented in (9) using domain notation. In the illustration, strand *C* binds to multi-stranded substrate complex *S* to form intermediate *I*1, which subsequently reacts with fuel strand *F* to release output product *OP*, side product *SP* and the catalyst *C*. In the absence of *C*, the reaction between *F* and *S* is very slow. Thus, input strand *C* dynamically amplifies output *OP*; this is roughly analogous to a NPN transistor where *F* and *S* together act as the collector, *C* as the base and *OP* as the emitter.

Crucial to the function of this catalytic reaction are toeholds, short single-stranded DNA (ssDNA) domains that initiate binding between DNA strands and complexes.

*To whom correspondence should be addressed. Tel: +1 626 235 0375; Fax: +1 626 584 0630; Email: dzhang@dna.caltech.edu
Correspondence may also be addressed to Erik Winfree. Tel: +1 626 395 6246; Fax: +1 626 584 0630; Email: winfree@caltech.edu.

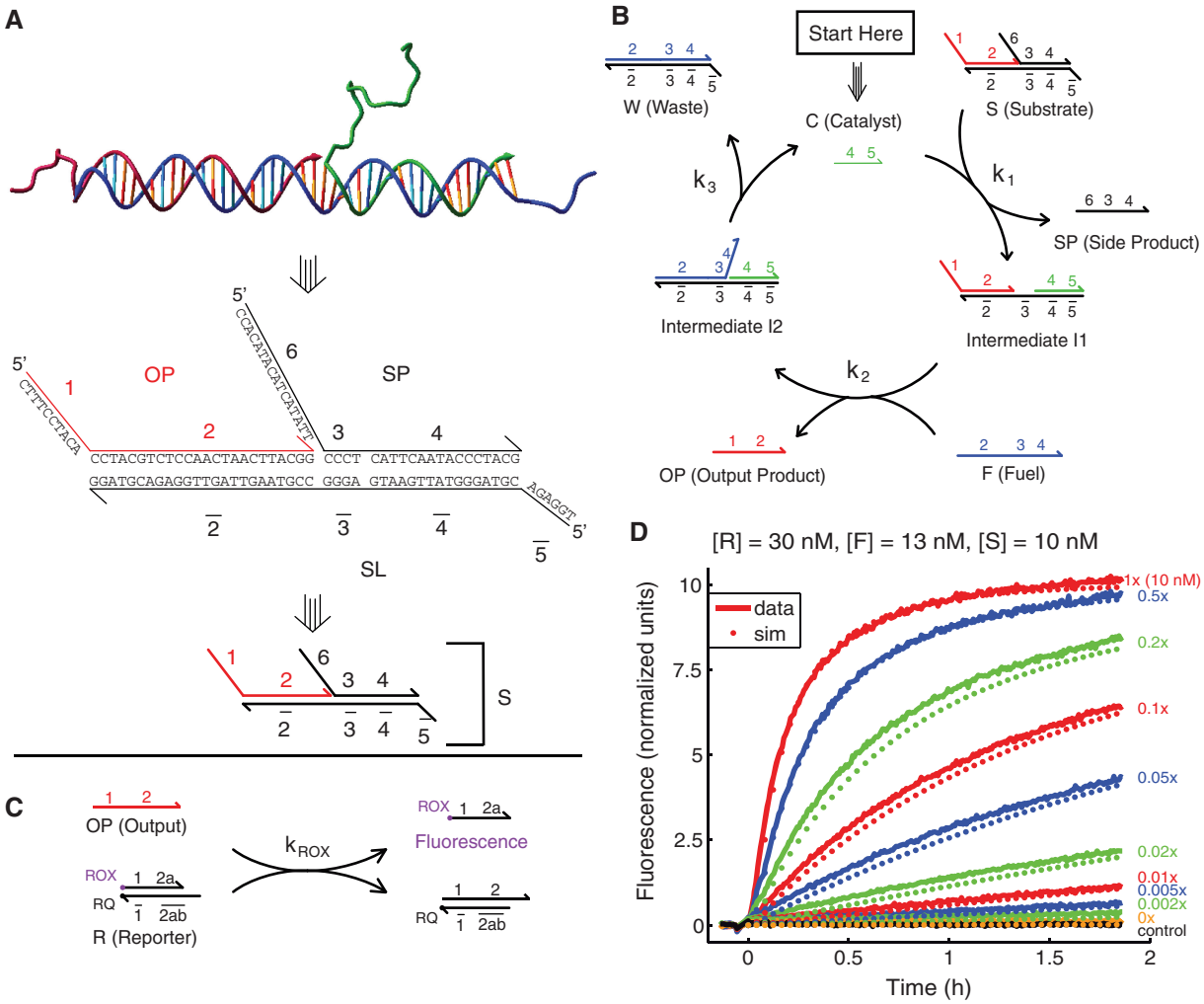

Figure 1. A non-covalent strand displacement reaction catalyzed a target ssDNA molecule *C* [adapted from (9)]. (**A**) DNA abstraction. The double-helix DNA molecule (top) is typically abstracted as two directional lines, one for each strand, with base identities shown (middle). Here, we abstract the DNA molecule one step further by grouping contiguous nucleotides into domains, functional regions of DNA that act as a unit in binding (bottom). Domains are labeled by numbers. Domain $\bar{x}$ is the complement of (and will hybridize to) domain x. The strands *OP*, *SP* and *SL* form the three-stranded DNA complex *S*. The DNA molecule in the top panel was drawn using Nanoengineer, a free DNA visualization software by Nanorex. (**B**) The designed mechanism of catalytic function. (**C**) Fluorescent reporter complex. Output product *OP* reacts stoichiometrically with reporter complex *R* to yield a fluorescent strand. ROX denotes the carboxy-X-rhodamine fluorophore (attached to the DNA molecule via an NHS ester), and RQ denotes the Iowa Black Red Quencher. This indirect reporter complex was used because of the thermodynamic effects of fluorophore-quencher binding (20). From (9), k_{ROX}, the second-order rate constant of reaction between *OP* and *R*, was measured to be 4×10^5 $M^{-1}s^{-1}$. In experiments, the concentration of the reporter *R* was in excess of the concentration of the fuel *F* and substrate *S* to minimize the reporter delay (no more than 2 min for $[R] = 30$ nM). (**D**) Experimental and simulation results from (9). Dotted lines show ordinary differential equation (ODE) simulation results according to the model in Table 2.

In Figure 1B, the reaction between catalyst *C* and substrate *S* is initiated by the hybridization of toehold domain 5 on *C* to its complement $\bar{5}$ on *S*. The binding energy of toehold domains have been shown to sensitively affect the kinetics of strand displacement reactions (2,3).

The catalyst *C* can potentially be any single-stranded oligonucleotide with limited secondary structure: the sequences of substrate *S* and fuel *F* are designed so that their reaction is catalyzed by strand *C* of a given sequence. Strand *C* is abstracted here as the concatenation of the two domains 4 and 5, such that the toehold domain 5 is the six 3′-most nucleotides, and domain 4 contains the remaining bases in *C*. The kinetics of this catalyst system are the fastest when *C* and *F* possesses no secondary structure; significant secondary structure in *C* or *F* will slow the kinetics of the reaction because *C* and/or *F* must spontaneously unfold before they can react with *S* and *I*1, respectively.

Previous work established that the reaction shown in Figure 1B is robust to the temperature and salt concentration changes (9). Here, we characterize other robustness and modularity properties of this molecular amplification mechanism, including:

- Maximum catalytic turnover: the maximum turnover is the average number of reactions catalyzed by each catalyst molecule, and determines the length of time that the system proceeds as designed. High turnover is necessary for sustained robust function.

- Effects of overhangs on the catalyst molecule: the catalyst molecule in our system is ~20 nt long. For amplifying and modularly interfacing with longer nucleic acids, unique subsequences could be used if overhangs do not significantly affect catalytic activity.
- Importance of the 5′/3′ orientation: proper function of the mechanism with inverted 5′/3′ orientation allows greater flexibility in interfacing with other nucleic acid systems, enhancing modularity.
- Sensitivity to mutations in the fuel and catalyst: high reaction specificity would ensure that only the desired molecule triggers the programmed reaction and allows for the simultaneous operation of many different modular subsystems.
- Effects of strand impurities: imperfect synthesis of oligonucleotides can lead to reactions other than those of the desired pathway. While impurities can be removed via purification, systems robust to strand impurities have a clear advantage.
- Interference from DNA molecules of randomized sequence: sensitivity to crosstalk from unrelated molecules would reduce modularity in engineered systems and undermine robust function in environments containing biological molecules.
- Sequence robustness of the design: stringent sequence design requirements would limit the number of DNA components that can simultaneously be in solution, and reduce the modularity and robustness of the design.

In essence, we are interested in characterizing the properties of this particular mechanism so that it can be used as a 'plug and play' module for incorporation into generalized chemical reaction networks. While the properties of other DNA strand displacement-based devices vary somewhat by design, it is hoped that many of these components possess similar robustness and modularity properties, so that the results presented in this work are generalizable.

MATERIALS AND METHODS

Buffer conditions

The buffer for all experiments was TE (10 mM Tris–HCl pH balanced to 8.0, 1 mM EDTA), purchased as 100× stock (Sigma-Aldrich), with 12.5 mM $MgCl_2$ added. Because EDTA chelates magnesium ions, the effective concentration of Mg^{2+} is 11.5 mM. All experiments and purifications were performed at 25°C.

DNA sequences and design

The sequences of the basic catalyst system shown in Figure 1B are the same as those in (9), which were carefully designed by hand to avoid secondary structure in single-stranded species (21) (Table 1). To this end, one useful sequence design heuristic was to minimize the number of G's in single-stranded species (e.g. *C*, *F* and *OP*). Not only are unintentional G–C bases much more stable than A–T ones, G's also further contribute to undesired secondary structures through G–T wobbles and G-quartets (24). For this reason, there are a total of only four G's in domains 1–6 (out of 76 bases). The sequences for the 5′/3′-inverted catalyst were similarly designed to minimize G content. Domains 10–13 possess only three G's (out of 40 bases).

Table 1. Domain sequences

Dom.	Sequence	Length (nt)
1	5′-CTTTCCTACA-3′	10
2a	5′-CCTACG-3′	6
2b	5′-TCTCCA-3′	6
2c	5′-ACTAACTTACGG-3′	12
2	5′-CCTACGTCTCCAACTAACTTACGG-3′	24
3	5′-CCCT-3′	4
4	5′-CATTCAATACCCTACG-3′	16
5	5′-TCTCCA-3′	6
6	5′-CCACATACATCATATT-3′	16
7	5′-TTCACCTCAGTTATG-3′	15
8	5′-TCAATTCCTAACATA-3′	15
9	5′-TTTTTTTTTTTTTTA-3′	15
10	5′-CACACA-3′	6
11	5′-ACTTCAGTCATTAAGC-3′	16
12	5′-AGAC-3′	4
13	5′-CCATACAAGTATCA-3′	14

Sequences for domains 1–6 are from (9).

The sequences used for the four-letter alphabet catalyst system shown in Figures 10 and 11 were designed by hand to possess minimal secondary structure while balancing the base composition of the domains (Table 5): domains 41–46 possess 16 G's, 16 C's, 26 A's and 21 T's. Domains 51 and 52 possess 7 G's, 10 C's, 14 A's and 6 T's. Domains 63–65 were also designed to minimize the number of G's, and together possess 1 G, 11 C's, 7 A's and 5 T's. Domains 71–76, being the 5′/3′ inverse of domains 1 through 6, have the same nucleotide distribution as the latter. The theoretical and experimental effectiveness of using three- and four-letter alphabets are discussed in the text.

Substrate purification

DNA oligonucleotides used in this study were purchased from Integrated DNA Technologies (IDT), purified by high-performance liquid chromotography (HPLC), except as noted in the text. Where applicable, fluorophores and quenchers were attached by IDT as well.

We further purified all multi-stranded complexes (i.e. substrates *S* and the reporter complexes *R*) by non-denaturing (ND) polyacrylamide gel electrophoresis (PAGE) as follows: strands for each sample were prepared at 20 μM concentration, subject to pipetting, dilution and extinction coefficient errors. The samples were then annealed with an Eppendorf Mastercycler Gradient thermocycler, cooling from 95°C to 20°C at a constant rate over the course of 90 min.

ND loading dye containing Xylene Cyanol FF (XCFF) in 50% glycerol was added to all samples, achieving final gycerol concentration of 10% by volume. The samples were run on 12% ND PAGE at 180 V for 6 h at 25°C (using a Novex chamber with external temperature bath).

The proper bands were cut out and eluted in 1 ml TE/Mg^{2+} buffer for 2 days. Typical yields ranged from 30% to 60%. Purified complexes were quantitated by measurement of absorbance at 260 nm using an Eppendorf Biophotometer, and concentrations were calculated using extinction coefficients for single- and double-stranded DNA (dsDNA) predicted by nearest neighbor models (22).

Denaturing PAGE

A 12% acrylamide (19 : 1 acrylamide: bis) denaturing PAGE was run to prepare Figure 8C. Acrylamide solution was diluted from 40% stock acrylamide (Ambion). Denaturing loading dye containing XCFF in 80% formamide was added to all samples in 1 : 1 ratio, achieving final formamide concentration of 40% by volume. Gels were run at 120 V for 1 h at 25°C, with temperature controlled using a Novex chamber external temperature bath. Gels were stained with Sybr-Gold stain (Invitrogen), and scanned with a Bio-Rad Molecular Imager.

Spectrofluorimetry studies

Spectrofluorimetry studies were done using a SPEX Fluorolog-3 (Horiba) with 1.6 ml synthetic quartz cells (Hellma catalog number 119-004F). Sample solutions were excited at 588 nm, and emission at 602 nm was observed (optimal signal for ROX fluorophore in our buffer). Slit size used were 2 nm for both excitation and emission monochromators. Fluorescence experiments were done with integration time of 10s for every 60 s time-point.

Prior to each experiment, all cuvettes were cleaned thoroughly: each cuvette was washed 15 times in distilled water, once in 70% ethanol, another 5 times in distilled water, and finally once more in 70% ethanol. Appropriate volumes of DNA stock solutions, typically 1 μM in concentration, were added to TE/Mg^{2+} buffer to achieve the correct final concentrations with a total volume of 1.5 ml in the cuvettes. Solutions were mixed by pipetting 250 μl of solution in and out of the cuvette 32 times. Stir bars were not used because they have been observed to significantly contribute to fluorophore decay. Unpublished data indicate that this may be due to bleach (used for cleaning) slowly desorbing from the surface of the stir bars, which in turn cause non-specific fluorophore quenching (B. Yurke, Personal communication).

For the slit size, concentrations and times chosen, no measurable photobleaching was observed. Fluorescence measurements are linear in the concentration of the free fluorescent strand *F*. All experimental results were within the linear regime of the spectrofluorimeter detector, according to specification sheets provided by the manufacturer.

Fluorescence normalization

Fluorescence is normalized so that 1 normalized unit (n.u.) of fluorescence corresponds to 1 nM of unquenched fluorophore-labeled strand *F*. This normalization is based on the fluorescence levels of annealed samples: a negative control with only $[R] = 30$ nM (normalized to 0 n.u.), and a positive control with $[R] = 30$ nM, $[F] = 20$ nM and $[S] = 10$ nM (normalized to 10 n.u.). Normalization experiments were run once for each different purified reporter complex. Day-to-day and sample-to-sample variations are estimated to be <5% (9).

Carrier strands

It has been observed that DNA sticks non-specifically to pipette tips, so that serial dilutions lead to stocks more dilute than expected (9). Unfortunately, this loss is not consistent, so we could not compensate for tip loss with additional reagent. Instead, we introduced into all dilute stocks (1 μM and below) a non-reactive 20 nt poly-T 'carrier' strand, at a concentration of 1 μM. Since pipette tip loss is non-specific, the majority of DNA loss would be of the carrier strand, and serially diluted stocks are only slightly more dilute than expected. Poly-T strands have minimal influence on the reactions of other DNA molecules in this system (9).

Parameter fitting

The best fit rate constants to the revised ODE model of the system (Table 3) were fitted using the 'fminunc' function in Matlab to minimize the error between experimental data and the reaction model. The error is calculated as follows:

$$\text{Error} = \sum_{t,\text{traces}} \frac{(F_d(t) - F_m(t))^2}{t_{\max}.[S]_0^2}$$

where $F_d(t)$ is the fluorescence value of the data at time t, and $F_m(t)$ is the fluorescence value predicted by the ODE model at time t. The denominator terms correspond to a normalization factor to ensure that each plot contributes roughly equally in fits: $t_{\max}$ denotes the endpoint time of each experiment, and $[S]_0$ denotes the initial concentration of the substrate S (the maximum fluorescence level of a catalytic reaction that achieves 100% completion). The rate constants' confidence intervals were determined as the values at which the error value would double, holding all other parameters constant.

RESULTS AND DISCUSSION

Turnover characterization and revised modeling

The *turnover* is the number of reactions catalyzed by each catalyst molecule over a particular period of time. We can calculate the turnover from our experimental data by dividing the excess number of product molecules (over the uncatalyzed reaction) by the number of catalyst molecules present (catalysis, like other molecular processes, is stochastic, so we are actually calculating the average turnover over all the catalyst molecules present). In a perfect catalyst system, the turnover indefinitely increases linearly with time. In reality, however, the turnover will asymptotically approach a maximum value as time goes to infinity, because the catalyst will be inactivated by side reactions. In this section, we seek to characterize the

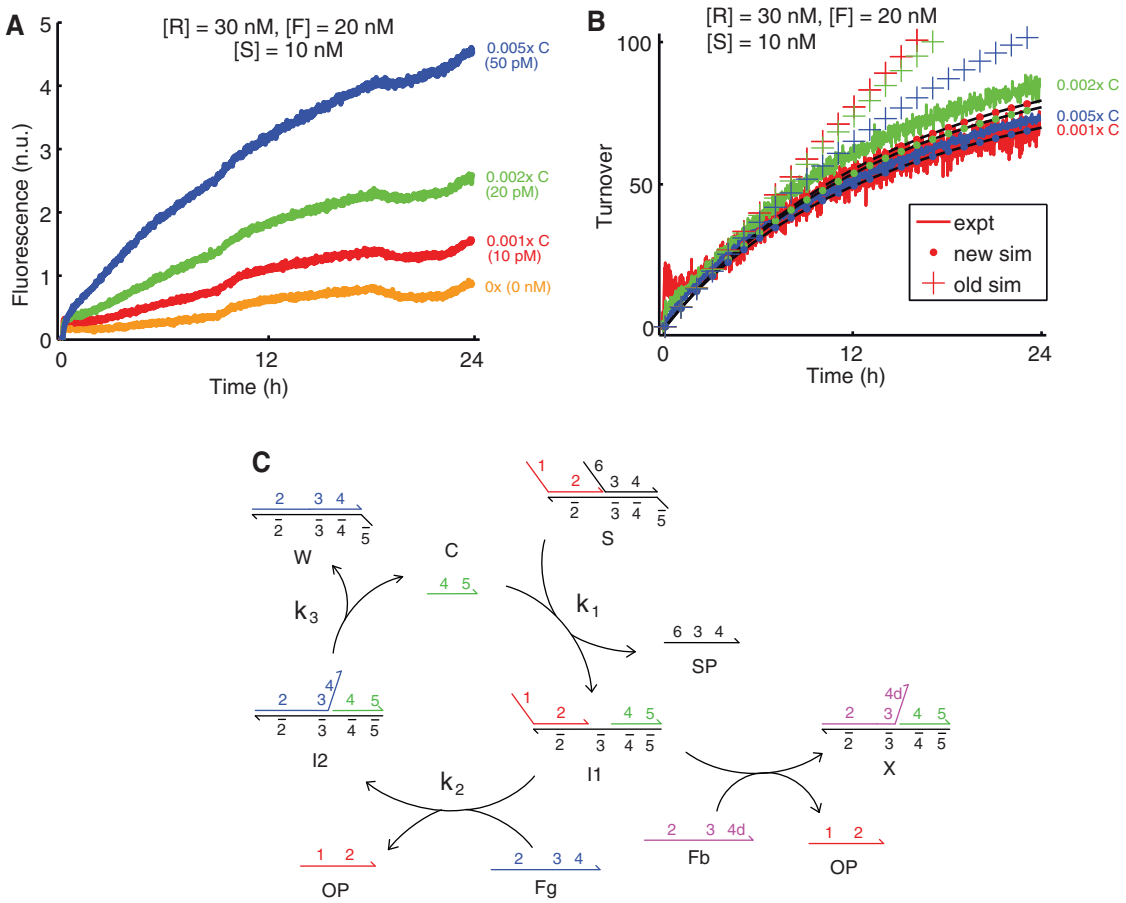

Figure 2. Catalytic turnover. (**A**) Raw data for turnover experiments. Traces showed significantly more noise than typical; possibly, this is due to lamp and temperature instability. (**B**) Turnover plotted as a function of time. Turnover is calculated as the excess normalized fluorescence above leak (0× trace) divided by concentration of catalyst; e.g. $T_{0.001x}(t) = F_{0.001x}(t) - F_{0x}(t)/0.001 \cdot 10\text{nM}$. 'Old sim' denotes simulations using the model presented in (9), which was fitted only to the 10 nM data shown in Figure 1D. The 'new sim' simulations use the model and rate constants fitted to the data presented in this article. The relative ordering and the quantitative differences between the 'new sim' simulations and the experimental results are not considered significant—all traces are considered to be within experimental error of one another. (**C**) Our new model that accounts for *F* subpopulations. A small fraction (fitted to be 1.0%) of *F* exists as *Fb*, with deletions in domain 4. These react with *I*1 to yield *X*, from which *C* cannot dissociate.

maximum turnover of the DNA catalyst system, and to explore potential causes for this limitation.

Figure 2A shows the catalytic activity using a high ratio of substrate to catalyst ([*S*]/[*C*] between 200 and 1000), and Figure 2B shows the inferred turnover. High substrate to catalyst ratios were used for this experiment so that turnover is not limited by substrate quantity. The catalytic turnover for all three catalyst concentrations approached a value between 80 and 100 after 24 h of reaction. Since the catalytic activity showed a consistent decreasing trend across all three experiments, it is likely that a real, but as yet undocumented, side reaction is inhibiting catalysis. The turnover after 24 h is significantly lower than 200, 500 and 1000 (the turnover at 100% reaction completion for 0.005×, 0.002×, and 0.001× catalyst), so substrate depletion is unlikely to be the main cause of this slowdown.

One likely cause of limited turnover is DNA strand impurities. For example, a single deletion near the 3′-end of the fuel strand *F* could slow the release of catalyst *C* from intermediate *I*2. With two or more deletions near the 3′-end of the fuel *F*, the release of *C* from *I*2 could be thermodynamically unfavorable. Studies on the heterogeneity of commercially synthesized oligonucleotides (25) indicate that DNA oligonucleotide samples show deletions at every single base.

To model fuel impurities and observe whether such impurities would produce simulation results consistent with experimental data, we modified our model to assume that the total *F* concentration is the sum of two distinct fuel molecules: a 'good' fuel *Fg* that reacts with *I*1 and release *C* from *I*2 as designed, and a 'bad' fuel *Fb* that reacts with *I*1 to form a new product *X* from which *C* cannot dissociate (Figure 2C). The relative concentrations of *Fb* and *Fg* are believed to depend on synthesis and purification of the *F* strand, and thus will likely differ from sample to sample of *F*.

Our new model fits five parameters: the rate constants k_0, k_1, k_2 and k_3, as well as $\chi = [Fb]/([Fb]+[Fg])$, the fraction of bad fuel molecules with deletions near the 3′-end. Best fit values under the new model are shown in Table 3. For comparison, the model presented in (9) is reproduced in Table 2. This new model shows a significantly better quality of fit to the data in Figures 2,

3 and 8, and implies a maximum achievable turnover of $\frac{1}{\chi} \approx 100 \pm 20$ for our current stock of *F*.

Figure 3 shows the behavior of the catalyst system with various fuel and substrate concentrations ranging from 1 nM to 100 nM. The catalytic behavior is qualitatively similar for all of the tested concentrations, and our new model quantitatively predicts the kinetics for all of these experiments. Thus, the catalyst system functions reliably and predictably over at least 2 orders of magnitude of concentrations.

Table 2. Original model

$S + C \underset{k_2}{\overset{k_1}{\rightleftharpoons}} I1 + SP$	$K_1 = 6.5\times10^5\ M^{-1}\ s^{-1}$
$I1 + F \xrightarrow{k_2} I2 + OP$	$K_2 = 4.2\times10^5\ M^{-1}\ s^{-1}$
$I2 \underset{k_1}{\overset{k_3}{\rightleftharpoons}} W + C$	$K_3 = 4\times10^{-3}\ s^{-1}$
$S + F \xrightarrow{k_0} OP + SP + W$	$K_0 = 2.3\times10^1\ M^{-1}\ s^{-1}$

Table 3. Revised model

$S + C \underset{k_2}{\overset{k_1}{\rightleftharpoons}} I1 + SP$	$k_1 = 2.7 \pm 0.5 \times 10^5\ M^{-1}\ s^{-1}$
$I1 + Fg \xrightarrow{k_2} I2 + OP$	$k_2 = 1.1 \pm 0.7 \times 10^6\ M^{-1}\ s^{-1}$
$I2 \underset{k_1}{\overset{k_3}{\rightleftharpoons}} W + C$	$k_3 = 1.1 \pm 0.5 \times 10^{-2}\ s^{-1}$
$S + Fg \xrightarrow{k_0} OP + SP + W$	$k_0 = 5\ M^{-1}\ s^{-1}$
$I1 + Fb \xrightarrow{k_2} X + OP$	
$S + Fb \xrightarrow{k_0} OP + SP + W2$	$\chi = \frac{[Fb]}{[Fg]+[Fb]} = 1.0 \pm 0.3 \times 10^{-2}$
$C + W2 \xrightarrow{k_1} X$	

Overhangs in the catalyst input

One long-term goal of strand displacement-based nucleic acid devices and circuits is to interface with biological systems, taking biological nucleic acid molecules as inputs and producing active biomolecules as outputs. When using long biological nucleic acids (such as most mRNAs) as inputs, it is often desirable to use only a unique unstructured subsequence as the input to the synthetic DNA devices to improve kinetics, specificity and cost. Consequently, the active input subsequence will usually possess 5′- and/or 3′-overhangs. Additionally, 5′- and 3′-overhangs may also be present in output signals of strand displacement-based DNA devices, as artifacts of upstream sources. For example, in (9), the output of the first catalytic reaction in the feed-forward network contained an extra 3′ domain that served no catalytic purpose in the downstream reaction. Here, we study the effects of these overhangs on the kinetics of the studied catalytic reaction.

We constructed five alternative versions of the catalyst with single- and double-stranded 5′- and 3′-overhangs (Figure 4A–E). The sequences of domains 7 and 8 were designed to be minimally interactive with other domains existing in the reaction, while domain 9 is mostly poly-T, and should exhibit the least interaction with other single-stranded domains (Table 1). Figure 4 shows the catalytic activities of these five modified catalysts. 3′-overhang, whether single- or double-stranded, have

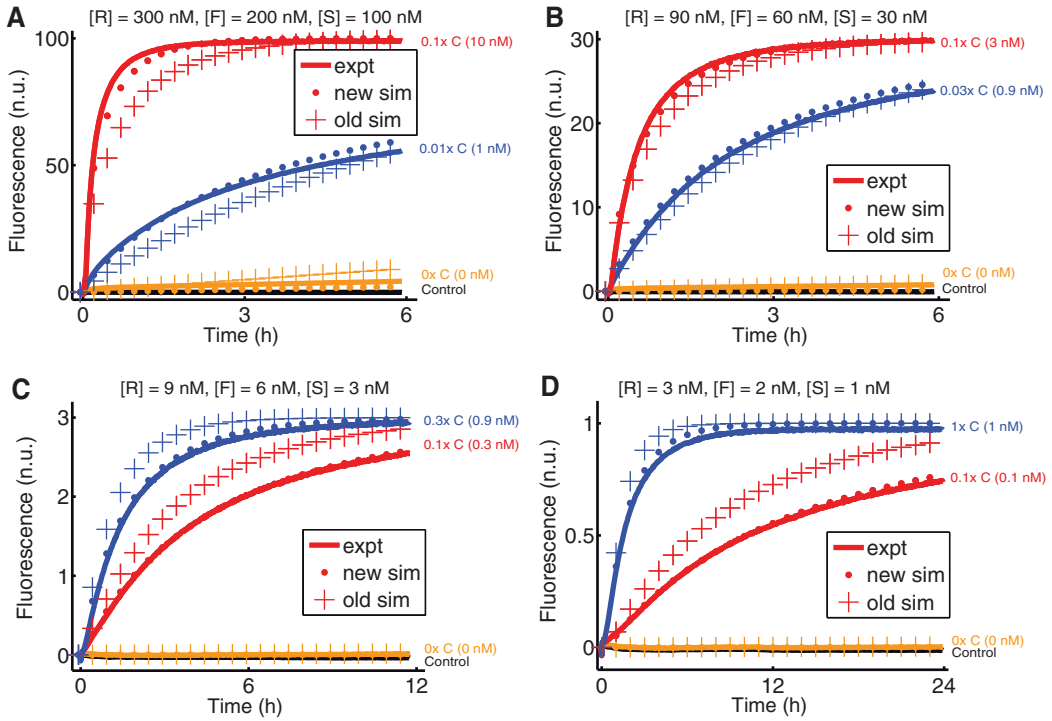

Figure 3. Catalytic function and model results for (**A**) 100 nM, (**B**) 30 nM, (**C**) 10 nM and (**D**) 1 nM substrate *S* concentration. The new model that accounts for catalyst inactivation (Table 3) fits the experimental data better than the old model (Table 2). Red traces denote 0.1× catalyst, whereas blue traces denote catalyst concentrations of ~1 nM.

relatively minimal effects on the catalytic activity. In contrast, 5′-overhangs significantly reduce catalytic activity: a double-stranded 5′-overhang reduces catalytic activity by a factor of roughly 2, and a single-stranded 5′-overhang reduces catalytic activity even further. The catalytic activity of the catalyst with a single-stranded domain 8 5′-overhang is roughly 1/8 that of the standard catalyst, or about a factor of 4 lower than a catalyst with a double-stranded domain 8 5′-overhang. In contrast, the catalyst with a single-stranded domain 9 5′-overhang exhibited catalytic activity roughly 1/2 that of the standard catalyst, and comparable with the catalyst with a double-stranded domain 8 5′-overhang.

One interpretation of these results is that the catalysts with 5′-overhangs cause a reduced rate of reaction between intermediate *I*1 and *F* by inhibiting the accessibility of domain $\bar{3}$ (see Figure 1B). A double-stranded 5′-overhang on *C* hinders *F* from binding due to electrostatic repulsion and steric hindrance; a single-stranded 5′-overhang on *C* additionally could fleetingly bind to the domain $\bar{3}$ via spurious Watson–Crick complementarities. For example, the domain 8 contains a subsequence 'CCT' could fleetingly bind to the 'AGG' subsequence of $\bar{3}$, despite the fact that NUPACK (27) does not predict this binding to be significant (folding at 25°C shows $\bar{3}$ to be unbound with probability greater than 0.999 in the modified *I*1 state with the single-stranded domain 8 5′-overhang). The extent to which spurious bindings can occur is the likely cause of the difference in catalytic activities between traces (D) and (E).

If this explanation is correct, then the effects of overhangs depend on the relative position of the overhang, with regard to the catalytic substrate *S*. In the current scheme, the catalyst's 5′ domain acts as the recognition sequence; consequently, 5′-overhangs affect the kinetics of the catalytic reaction. In a scheme where the catalyst's 5′ domain acts as the toehold and the 3′ domain acts as the recognition sequence, presumably 3′-overhangs on the catalyst would slow the kinetics of catalysis and 5′ overhangs would have minimal effect on kinetics.

Although the kinetic slowdown due to 5′-overhangs is unfortunate, it is encouraging that the catalytic reaction qualitatively works similarly even when the catalyst possesses 5′- and 3′-overhangs. This implies that such systems can be used in conjunction with biological nucleic acid inputs.

5′/3′ orientation

The dominant catalytic pathway presented in Figure 1B was designed to function based on the principles of toehold exchange, which have been characterized in (2). Thus, it is expected that the same mechanism would function in a design in which all 5′/3′ orientations were inverted (Figure 5A). A catalytic system with an inverted

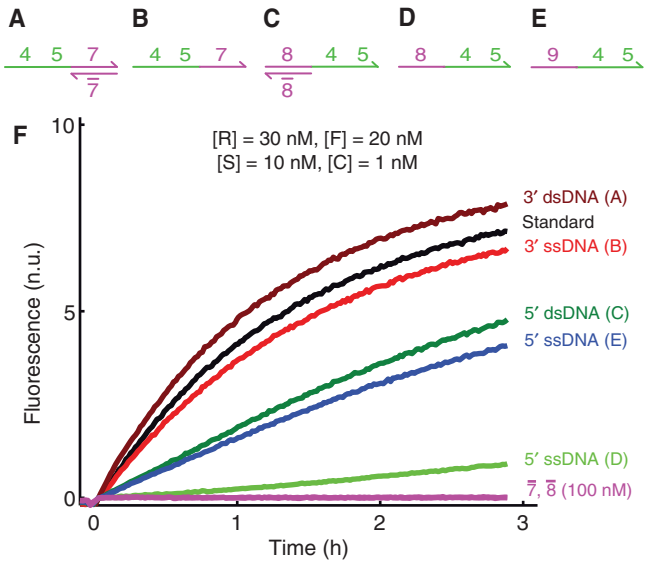

Figure 4. Effects of overhangs on catalytic activity. Schematics for catalyst *C* with a (**A**) 3′ dsDNA overhang, (**B**) 3′ssDNA overhang, (**C**) 5′dsDNA overhang, (**D**) 5′ssDNA overhang and (**E**) 5′ssDNA overhang (mostly poly-T). (**F**) Catalytic activity of catalyst molecules with various overhangs. The magenta traces show that the $\bar{7}$ and $\bar{8}$ strands possess no catalytic activity on their own, and that the increased catalytic activity of (A) and (C) over (B) and (D) is only due to the single/double-stranded state of the overhang.

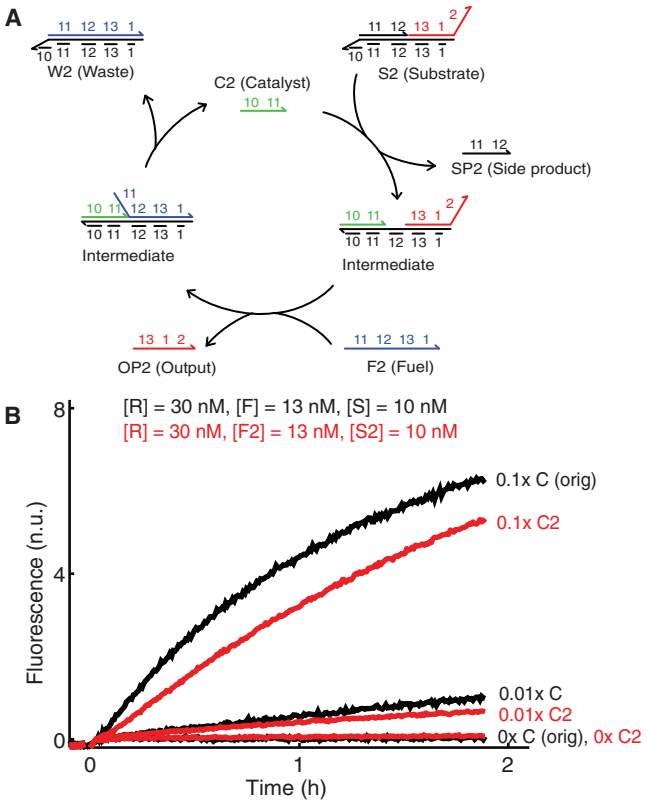

Figure 5. 5′/3′ inverted catalyst and substrate. (**A**) Schematic and (**B**) results. This system uses the same reporter complex as the original catalytic reaction (shown in Figure 1C). Substrate *S*2 does not react significantly with reporter complex *R*. Note that the single-stranded domain 2 on *S*2 could hybridize to the complementary domain $\bar{2}$b on reporter *R*. From there, it is possible to initiate a four-way branch migration process that could result in the release of the fluorophore-labeled DNA. However, four-way branch migration processes are significantly slower than three-stranded branch migration, and the hybridization of domain 2b is transient enough that this unintended pathway does not seem to be significant at our experimental conditions.

5′/3′ orientation may be desirable to improve kinetics in cases where the catalyst possesses a 5′-overhang but not a 3′-overhang, and in certain cases to prevent spurious interactions with other strands existing in solution. Experimentally, the catalytic system with inverted 5′/3′ orientation functions qualitatively and quantitatively similar to that of the original reaction (Figure 5B). In these experiments, all DNA strands used were HPLC purified. See also 'Effects of impurity and 5′/3′ orientation on maximum turnover' section.

Catalyst specificity

Regardless of whether reactions take place in a biological cell or in a test tube, in all but the simplest reaction networks there will exist a large number of molecules that may interfere with programmed reactions via spurious binding interactions. One hallmark of a well-designed molecular amplifier is specificity: only the exactly correct input should trigger amplification. In the case of chemical amplification through DNA catalysis, specificity can be quantitatively measured as the amount of catalytic activity caused by oligonucleotides differing in sequence from the designed catalyst.

As the most stringent measure of specificity, we characterize the catalytic activities of oligonucleotides differing from the proper catalyst by only a single nucleotide. Insertions, deletions and point 'mutations' were introduced at positions 3, 7, 12 and 17 on the catalyst strand, numbered from the 5′ end. Figure 6A–C shows the catalytic activities of these mutant catalysts. Mutations showed strong suppression of catalytic activity, regardless of position. Insertions to the 5′ of positions 3 and 7 affected the catalyst's activity less than other tested mutations; the reason for this is not yet understood.

The suppression ratio plotted in Figure 6D summarizes the effects of the tested mutations on catalytic activity. The suppression ratio is calculated by dividing the initial fluorescence slope of standard reaction (linear fit to $0 < t < 0.5$ h data) with the fluorescence slopes of the mutant catalysts (linear fit to $0 < t < 3$ h data). The estimation of the initial fluorescence slope of the standard catalyst differs from those of the mutant catalysts because at $t > 0.5$ h, the standard catalytic reaction is

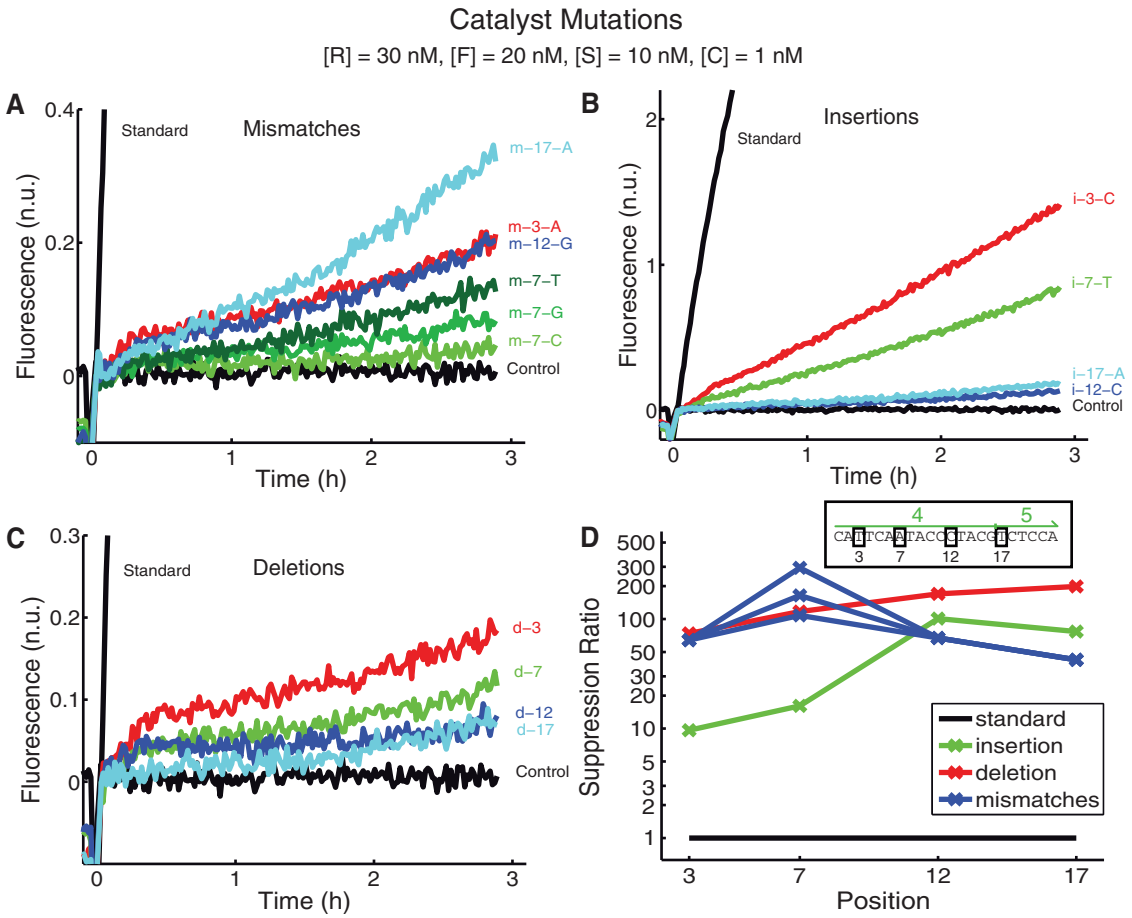

Figure 6. Effects of single base catalyst *C* mutations. (**A**) Mismatches. The trajectory labels show the position and new identity of the mutated base. For example, 'm-17-A' denotes that the 17th base of catalyst *C* (from the 5′-end), was mutated from thymine to a adenine. (**B**) Insertions. The inserted base is inserted before (5′ of) the position denoted. (**C**) Deletions. (**D**) Summary of suppression by various catalyst mutations. The suppression factor is calculated as the initial slope of activity by the standard trace divided by the initial slope of activity by the mutated catalysts: $S.F. = [F_s(10\text{min}) - F_s(0\text{min}))/(10\text{min})]/[(F_m(3\text{h}) - F_m(0\text{h}))/(3\text{h})]$, where $F_m(t)$ is the fluorescence value due to a mutated catalyst at time t. (D.inset) The sequence of the catalyst molecule *C* is shown with the positions at which mutations were performed.

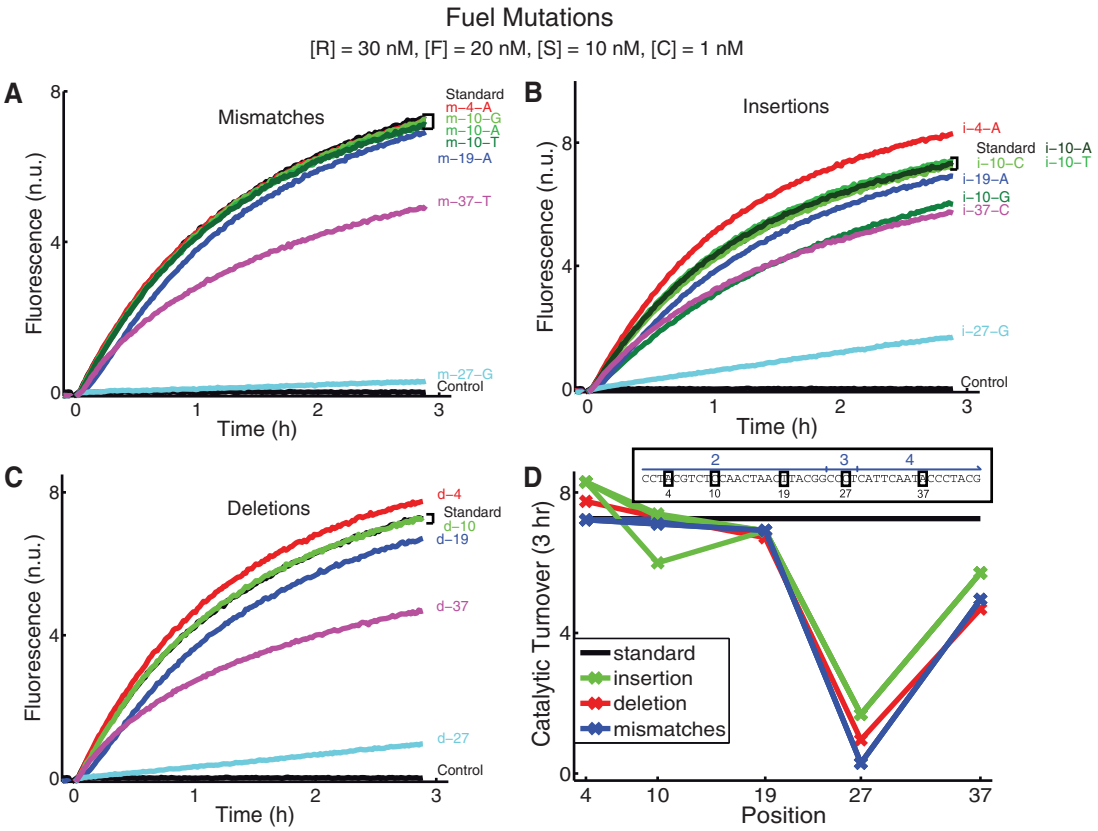

Figure 7. Effects of single-base fuel F mutations. (**A**) Mismatches. (**B**) Insertions. The inserted base is inserted before (5′ of) the position denoted. (**C**) Deletions. (**D**) Summary of the catalytic activity using mutant fuel molecules. (D.inset) Sequence of the fuel molecule F and positions of mutations.

noticeably slowed by substrate depletion and waste buildup.

All single-base mutations tested suppressed catalytic activity by at least a factor of 10, and the majority of them suppressed catalytic activity by $\geq$50-fold. This implies that in chemical reaction networks, DNA sequences can be easily designed to avoid catalytic crosstalk.

Fuel specificity

A similar study was performed on the effects of single-base mutations on the fuel; these results are plotted in Figure 7A–C. Unlike with catalysts, single-base mutations to the fuel strand yielded relatively little change in catalytic activity, except at position 27 and to a lesser extent position 37. Position 27 is in domain 3, which is used by the fuel F to bind intermediate $I1$; this explains why catalytic activity is sensitive to the bases in this domain. The weaker effect of fuel mutations at position 37 can likely be attributed to slower release of the catalyst from intermediate $I2$, because the branch migration process stalls at the fuel mismatch position.

The difference in specificity between the fuel and the catalyst is not accidental: the binding reaction between C and S to yield $I1$ and SP was designed to have $\Delta G^{\circ} \approx 0$ (Figure 1B). Thus, a mutation in C destabilizes intermediate $I1$ and raises the standard free energy of the reaction as well as the standard free energies of many undocumented intermediates of the reaction. Consequently, the activation energy is raised and the kinetics are slowed. In contrast, the reaction between F and $I1$ is thermodynamically favorable (making a helix in domain 3), so a single-base mutation to F does not greatly affect the spontaneity of the $F + I1 \underset{I}{\rightarrow} 2 + OP$ or the $I2 \rightarrow W + C$ reactions.

One alternative hypothesis for explaining the difference in sensitivity between catalyst and fuel mutations is that the proposed pathway shown in Figure 1B is only one of two predominant pathways—that is, $F + I1$ may yield $C + I3$, where $I3$ is a new intermediate complex composed of F, OP and SL. Thus, a single mutation in the fuel at position 19 would bias against the shown $F + I1 \underset{I}{\rightarrow} 2 + OP$, but not affect the $F + I1 \underset{I}{\rightarrow} 3 + C$ pathway, and the second pathway would become the dominant catalytic pathway. Intermediate $I3$ would subsequently release product OP and waste W. A single mutation in the fuel at position 37 would bias against the alternative pathway, but not affect the shown $F + I1 \underset{I}{\rightarrow} 2 + OP$ pathway.

Based on our previous work on the kinetics of toehold exchange (2), we do not expect that the $F + I1 \underset{I}{\rightarrow} 3 + C$ pathway is significant: the rate constant of a toehold exchange reaction with a 4-nt invading toehold and a 6-nt incumbent toehold is at least 2 orders of magnitude slower than that with a 0 nt incumbent toehold. Thus, we

expect that a catalytic reaction using a fuel with two mutations, one at position 19 and one at position 37, would have approximately the same activity as that of the fuel with a single mutation at position 37.

The simultaneous sensitivity of catalytic activity to catalyst mutations and the robustness to fuel mutations suggest the possibility of a universal fuel molecule. By designing catalysts that differ from one another by only a few bases, it could be possible to maintain reaction specificity while using the same fuel molecule to power multiple different catalytic reactions.

Strand purity effects

In working with DNA, a large fraction of the time and cost lies not in strand synthesis, but rather in strand purification. Modern DNA strand synthesis involves a large number of protection and deprotection reactions, which lead to inevitable errors in synthesis, in the form of various truncations, deletions and depurinations in the synthesized oligonucleotides. While oligonucleotides with synthesis errors can be mostly removed by post-synthesis purification procedures, such as PAGE and reverse-phase HPLC, some fraction of the purified oligonucleotides still inevitably contain deletion and truncation products (25). Furthermore, PAGE and HPLC strand purifications are expensive; the development of DNA devices robust to the oligonucleotides with synthesis errors would facilitate the widespread and practical adoption of nucleic acid-based synthetic biology.

Here, we observe the effects of using unpurified DNA oligonucleotides that contain a significant fraction of truncation products. The catalytic reaction of interest possesses two reactants (*S* and *F*), and one catalyst molecule (*C*). In experiments described in previous sections, all oligonucleotides involved were purchased with reverse-phase HPLC purification (by IDT). We tested the effects of replacing each of these with its unpurified counterpart, and characterized the extent to which the catalytic reaction tolerates synthesis impurities. Unpurified versions of the fuel, substrate and catalyst are denoted by *f*, *s* and *c*. Note that for *s*, although the three components of DNA strands *OP*, *SP* and *SL* are individually not purified, *s* is still manually purified by ND PAGE to ensure proper stoichiometry and complex formation (see 'Materials and Methods' section).

Figure 8A shows the behavior of the catalytic reaction when substituting unpurified strands. The kinetics of reactions involving unpurified strands were slower than the corresponding reaction using purified strands. Furthermore, the slowdown due to using unpurified strands was cumulative—for example, using unpurified fuel and unpurified catalyst yielded slower reactions than unpurified fuel (and purified catalyst and substrate), which was in turn slower than the standard reaction with purified fuel, catalyst and substrate. This is consistent with our results from earlier: it was shown in the previous sections that even single-base deletions can significantly impede the designed catalytic pathway. Thus, in using unpurified strands, the effective concentrations of active species are lowered.

The use of unpurified strands for fuel and substrate does not significantly change the rate constant of the uncatalyzed (leak) reaction (Figure 8B). Thus, the ratio of the catalyzed to uncatalyzed reaction rates is lower when using unpurified strands.

The degree of strand purity for both HPLC-purified and unpurified oligonucleotides are characterized by gel in Figure 8C and quantitated in Figure 8D–F. From the data in Figure 8D–F, the purities are calculated and summarized in Table 4.

The catalytic reactions performed qualitatively similarly and quantitatively slower by no more than a factor of 2 when using unpurified DNA strands, as compared with purified strands. This robustness result implies that faulty strands can still yield robust components for DNA strand displacement-based systems.

Robustness to background nucleic acid molecules

In order for the synthetic catalysts characterized here to be easily incorporated into a complex biochemical network, it must function in an environment rife with other molecules. These 'noise' molecules may interact non-specifically with the reaction substrates, catalysts or products, changing the kinetics and thermodynamics of the intended reactions. For example, in a cell, noise molecules include proteins, nucleic acids and small organic molecules. In (9), an autocatalytic variant of the catalytic reaction showed slower kinetics and higher leakage in the presence of cell lysate and total RNA.

In this section, we characterize the robustness properties of the catalytic reaction to DNA noise molecules. First, we tested the effects of a random mix of poly-N 50-mers (with G, C, A or T approximately equally likely at every base position). As shown in Figure 9, these random oligonucleotides never induced a false positive, in terms of catalyzing the release of product in the absence of the correct catalyst *C*. However, increasing concentrations of poly-N noise molecules did increasingly suppress the kinetics of the catalytic behavior; at 1 μM poly-N noise (100× that of the catalytic substrate), the catalytic activity is reduced by a factor of about 20. Contrast this result to results presented in (4,9), in which the introduction of substantial quantities of total RNA yielded relatively minor effects on the kinetics of strand displacement-based logic gates and catalytic systems. The poly-N oligonucleotides likely interact with stronger affinity for the single-stranded fuel *F* and catalyst *C* strands—in this case, the intrinsic secondary structures in biological mRNA molecules likely reduce nonspecific binding, especially because total RNA is mostly ribosomal RNA with defined structures.

For applications in synthetic chemical networks, it is desirable to possess both fast kinetics and minimal nonspecific interactions. As mentioned in the 'Materials and Methods', one strategy used was to minimize the frequency of G's in single-stranded domains and strands. DNA strands that do not include the G nucleotide are also unlikely to bind non-specifically to the fuel and catalyst molecules. This is consistent with experimental data (Figure 9): poly-H noise molecules (in which every

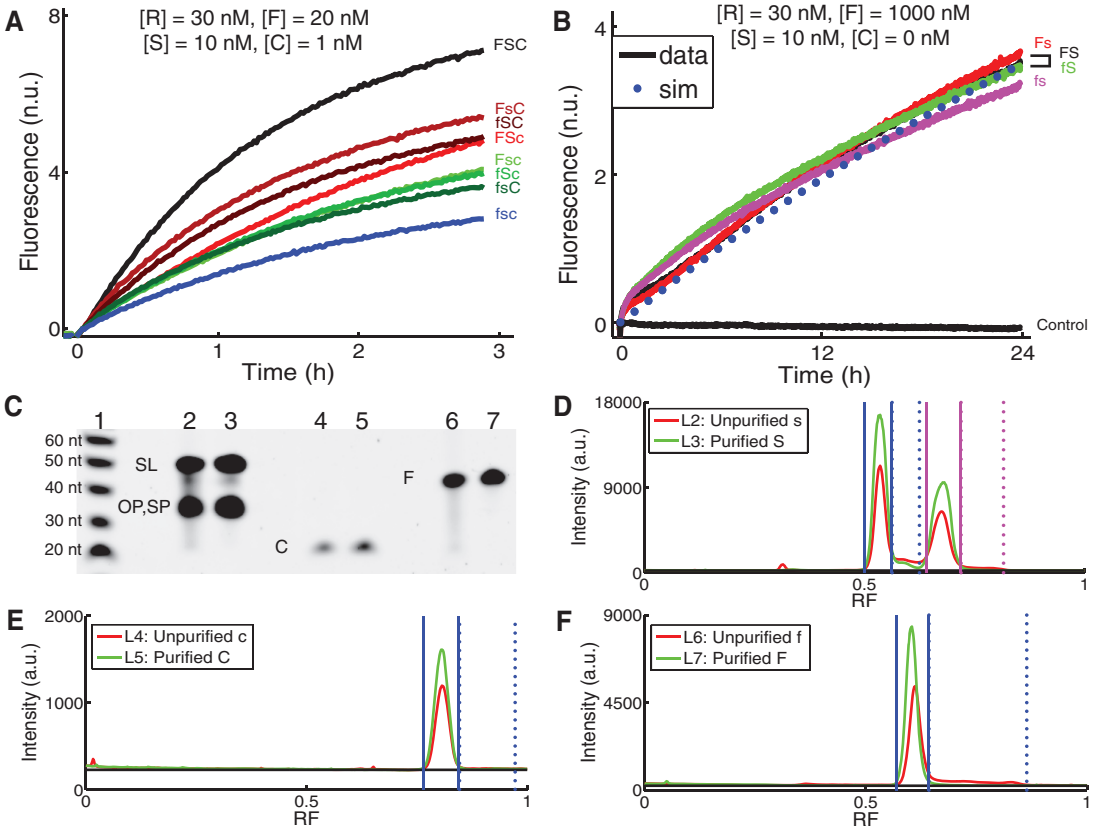

Figure 8. Behavior of the catalytic reaction using fuel, substrate and catalyst oligonucleotides with no post-synthesis strand purification. (**A**) Effects of using unpurified DNA on catalytic activity. Uppercase 'F' denotes fuel *F* purified commercially by HPLC, while lowercase 'f' denotes unpurified fuel. Similarly, uppercase 'S' and 'C' denote that the strands in *S* and the catalyst *C* were purified. Note that though the substrate *s* used unpurified DNA strands, it was still manually purified by PAGE to ensure correct stoichiometry. (**B**) Effects of unpurified DNA on the uncatalyzed reaction rate. The dotted rate shows simulation results for $k_0 = 5\ M^{-1}\ s^{-1}$. (**C**) Denaturing gel of purified and unpurified strands and complexes. Lanes 2 and 3 show the PAGE-purified substrate prepared from unpurified and purified *OP*, *SP* and *SL* strands, respectively. This shows the degree to which truncated strands are present in PAGE-purified *S* and *s*. Lanes 4 and 5 show the unpurified and purified catalyst, respectively. Lanes 6 and 7 show the unpurified and purified fuel, respectively. (**D**) Substrate *S* purity. Gel band intensities are displayed in arbitrary units (a.u.). Solid blue lines denote the (arbitrarily chosen) limits of the correct-length *SL* strand, while dotted blue lines denote the (arbitrarily chosen) limits of the truncated *SL*. The black horizontal line denotes the background intensity of the gel. The integrated intensity of the correct-length *SL* above the background is divided by the sum of it and the integrated intensity of truncated *SL* to yield the correct length fraction (summarized in Table 4). Similarly, solid magenta lines denote the limits of correct-length *OP* and *SP*, while dotted magenta lines denote the limits of truncated *OP* and *SP*. (**E**) Catalyst *C* purity. (**F**) Fuel *F* purity.

Table 4. Summary of correct-length fraction

Strand	Unpurified (%)	Purified (%)
SL	82	92
OP + SP	86	94
C	95	97
F	78	98

base is randomly C, A or T) had no significant effect on the catalytic reaction, even at micromolar concentrations.

Characterization of a catalyst system using a four-letter alphabet

Previous experiments on the catalytic DNA reaction shown in Figure 1 used domain sequences that were designed to intentionally minimize the frequency of G's in single-stranded domains. This strategy was meant to minimize spurious binding interactions between single-stranded domains, and is effective in the design of synthetic reaction networks with minimal crosstalk interference. However, this strategy cannot be applied for systems interfacing with biology, because naturally occurring nucleic acids possess a more uniform mixture of all four nucleotides.

To investigate whether quantitatively similar catalytic reactions could be designed using a roughly equal distribution of all four nucleotides, we designed the catalytic system shown in Figure 10A. The domain sequences were designed to possess minimal unwanted secondary structure, and are shown in Table 5. Mfold (23) predicts these sequences to possess similarly little secondary structure as the corresponding ones from the original system: at 25°C and 11.5 mM Mg^{2+}, the predicted standard free energy (ΔG°) of the minimum free energy (mfe) structure of *C*4 was +0.12 kcal/mol (compared with +0.28 kcal/mol for *C*), of *F*4 was −1.34 kcal/mol (compared with −1.79 kcal/mol for *F*) and of *OP*4 was −0.59 kcal/mol

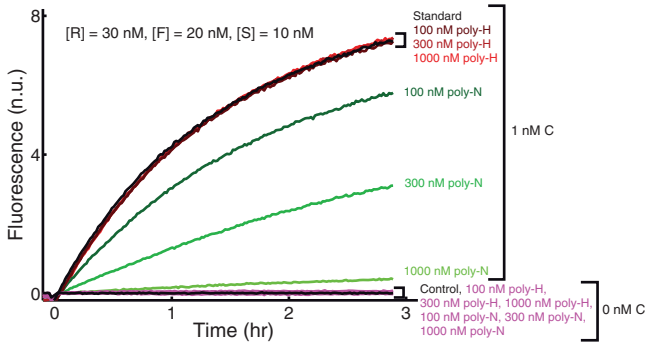

Figure 9. The effects of DNA molecular noise on the kinetics of the studied catalytic reaction. Poly-N and poly-H DNA noise molecules were employed. Poly-N molecules are a mixture of different oligonucleotides, each 50 nt long, and with each base in each oligonucleotide being randomly G, C, A, or T. Poly-H molecules are similarly a mixture of 44 nt oligonucleotides, in which each base is randomly C, A, or T.

(compared with −1.79 kcal/mol for *OP*) (note that the free energy of the completely unstructured state of the DNA is defined to be 0. The mfe values reported here are as calculated by mFold (23), and correspond to the free energy of the most stable structured state, which for certain structures is less stable than unstructured).

Experimentally, however, both the reporter and the catalytic reactions were significantly slower than their mostly three-letter counterparts (Figure 10B and C). Fluorescence characterization of the reporter complex kinetics showed a best-fit rate constant of reaction between *OP*4 and *OR*4 of $k_{R4} = 8.2 \times 10^4\ M^{-1}\ s^{-1}$. This is about a factor of 5 slower than the corresponding reporter complex designed using a three-letter alphabet (shown in Figure 1C). Similarly, the full four-letter system also exhibited significantly slower catalytic kinetics than the original catalytic system (shown in

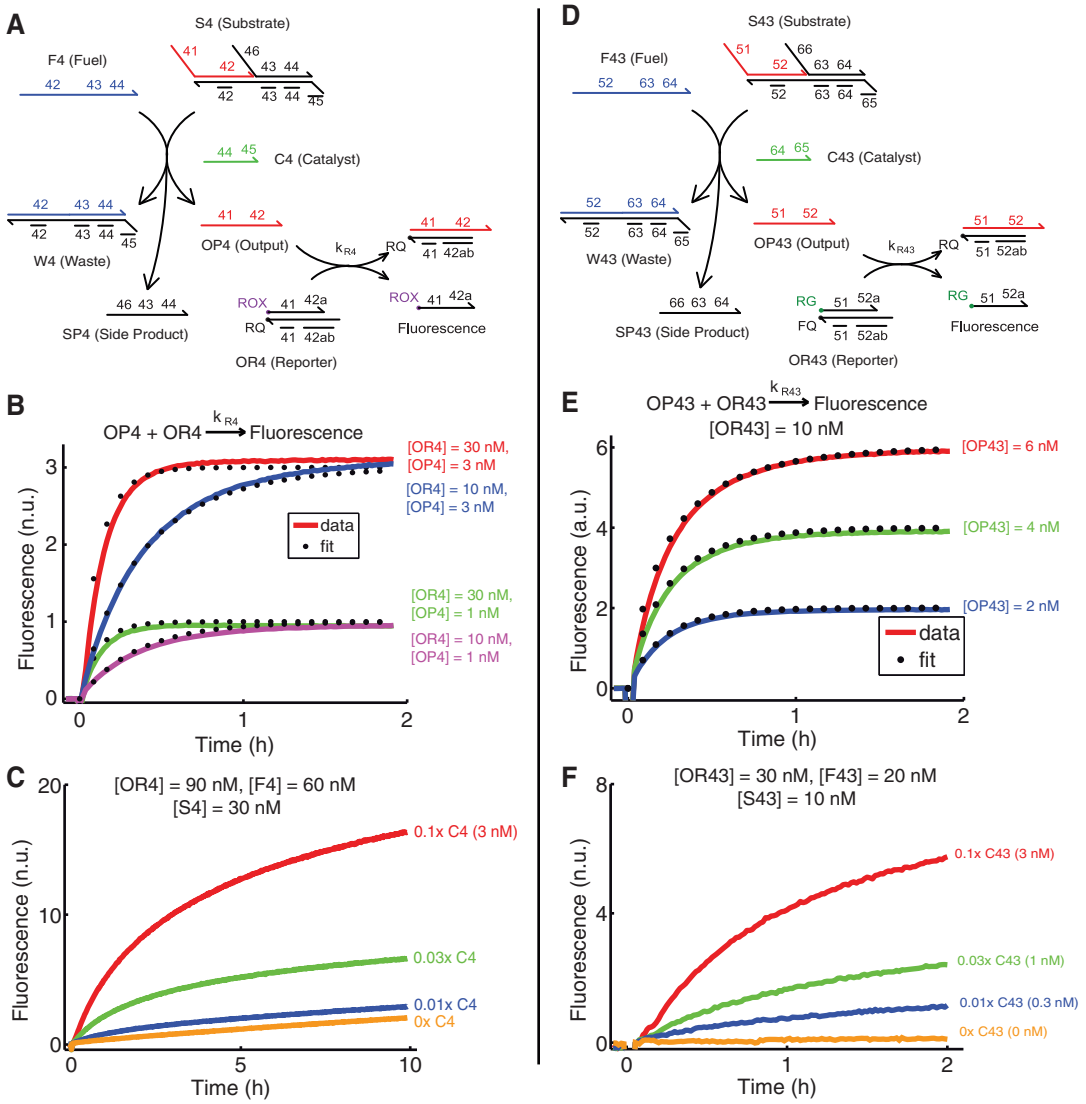

Figure 10. Catalytic design using four-letter alphabets. (**A**) Schematic of catalytic design using four-letter base sequences. Domain sequences are given in Table 5. (**B**) Fluorescence characterization of the kinetics of the reporter complex. The best fit value of k_{R4} (that produced the simulation traces shown as dotted lines) was $8.2\times10^4\ M^{-1}\ s^{-1}$. (**C**) Fluorescence characterization of the kinetics of the four-letter catalyst system. (**D**) Schematic of hybrid catalytic design using four-letter base sequences in the output domains 51 and 52, and a three-letter base sequences in the catalytic domains 63–66. Domain sequences are given in Table 5. 'RG' on the reporter complex denotes Rhodamine Green. 'FQ' on the reporter complex denotes the proprietary Iowa Black Fluorescence Quencher. (**E**) Fluorescence characterization of the kinetics of the reporter complex. The best fit value of k_{R43} (that produced the simulation traces shown as dotted lines) was $1.4\times10^5\ M^{-1}\ s^{-1}$. (**F**) Fluorescence characterization of the kinetics of the hybrid sequence catalyst system.

Table 5. Domain sequences for four-letter designs

Dom.	Sequence	Length (nt)
41	5′-TGTTACTGGCTCTGAT-3′	16
42a	5′-GACC-3′	4
42b	5′-AATGAAT-3′	7
42c	5′-ACCCGTTAC-3′	9
42	5′-GACCAATGAATACCCGTTAC-3′	20
43	5′-GAAAG-3′	5
44	5′-GCACTAAAAGTCTAC-3′	15
45	5′-CAATGTTC-3′	8
46	5′-ATAGAACATGTAGGT-3′	15
51	5′-ATAGATCCTGATAGC-3′	15
52a	5′-GAGAC-3′	5
52b	5′-CTAGCAA-3′	7
52c	5′-CCTGAAACCA-3′	10
52	5′-GAGACCTAGCAACCTGAAACCA-3′	22
63	5′-CCCTC-3′	5
64	5′-ATACAATACCC-3′	11
65	5′-TCACCATG-3′	8
66	5′-TTTTTTTTTT-3′	10

Figure 1B): 0.1× catalyst *C*4 took 10 h to achieve 60% reaction completion with [*S*4] = 30 nM, compared with 0.1× catalyst *C* taking 2 h to achieve 60% completion with one-third the substrate concentration ([*S*] = 10 nM). This implies the kinetics of the four-letter catalyst system to be roughly 15 times slower than the original catalyst system.

It was unclear why this four-letter system performed so much slower than the original system. To continue investigating the effects of the four-letter sequence designs, we next designed the 'hybrid' catalytic system shown in Figure 10D, in which the catalytic domains 63, 64 and 65 were composed of only C, A and T, respectively, while the output domains 51 and 52 contained all four nucleotides. At 25°C and 11.5 mM Mg^{2+}, the predicted standard free energy of the mfe structure of *F*43, *OP*43 and *C*43 were +0.02, −1.06 and +0.81 kcal/mol, respectively—this is similar to the values for the other systems. The characterization of the reporter complex for this system is shown in Figure 10E. The best fit rate constant for k_{R43} was 1.4×10^5 M^{-1} s^{-1}, which is slightly faster than that of the reporter shown in Figure 10, but significantly slower than the reporter on the original three-letter alphabet system.

The kinetics of this catalytic system (Figure 10F), however, are comparable with that of the original catalyst system (and thus 15 times faster than the four-letter system): at 10 nM substrate *S*43 concentration, 0.1× catalyst *C*43 (1 nM) achieved 60% reaction completion in $t \approx 2$ h.

The combination of the results shown in Figure 10 implies that usage of a four-letter alphabet causes a slowdown in the kinetics of toehold-mediated strand displacement and toehold exchange: both output products *OP*4 and *OP*43 are composed of all four nucleotides and the reaction between them and their respective reporter complexes were slow.

Finally, note that the uncatalyzed reaction was significantly faster for the four-letter system (Figure 10A–C) than the hybrid system (Figure 10D–F). We hypothesize that this is due to non-specific binding between the fuel *F*4 and single-stranded domain 46 on the substrate *S*4. This transient interaction results in an increased local concentration of *F*4 near *S*4, which in turn leads to a higher uncatalyzed reaction rate. In the hybrid system, the 66 domain is poly-T and unlikely to interact with *F*43.

Effects of impurity and 5′/3′ orientation on maximum turnover

Modern DNA synthesis proceeds from the 3′-end to the 5′-end. One effect of this asymmetry is that truncations and deletions near the 5′-end of DNA oligonucleotides are relatively more common than near the 3′-end. This, in turn, may cause catalytic systems with different 5′/3′ orientations to possess different properties. These effects will be particularly prominent when using strands not purified by HPLC or PAGE, due to the relatively higher fraction of truncated oligonucleotides.

One particular concern regarding the use of unpurified oligonucleotides is the effect of poorly synthesized fuel, even if it only exists in small quantities. For example, in the earlier section on modeling the turnover, it was hypothesized that limited turnover results from a small fraction of *F* possessing defects near the 3′-end. Small truncations near the 5′-end of the fuel *F* are unlikely to have a significant effect on the catalysis of the original system (9), as represented in Figure 11A. In contrast, in a catalytic system with inverted 5′/3′ orientation, 5′ truncations of the fuel could drastically reduce turnover (Figure 11B). For example, a fuel with a 4 nt truncation at the 5′-end (*Fmt* in the Figure) requires that the catalyst (*Cm*) must spontaneously dissociate 10 bp, rather than 6 bp, taking hours instead of seconds.

To experimentally investigate the effects of 5′/3′ orientation and purity on turnover with minimal confounding influence from sequence variation, we constructed a catalytic system using sequences that exactly mirrored the original catalytic system, with inverted 5′/3′ orientation. Unlike the 5′/3′ inverted system explored in Figure 5, this mirrored system will possess the exact same base distribution as well as the same nearest-neighbor distribution to the original system, and is assumed to have similar thermodynamic properties as the original system, when all oligonucleotides involved are perfectly synthesized. These mirrored domains are labeled with an '*m*' following the original domain/strand label (e.g. *Sm* and 1*m*, see Figure 11B and Table 6).

Because of the relatively higher frequency of truncations and deletions near the 5′-end of any synthesized oligonucleotide, it is expected that the mirrored system would thus yield lower maximum turnover than the original system when unpurified strands are used. Because oligonucleotides purified via HPLC or dual HPLC/PAGE likely still contain fuel strands with truncations and deletions, it is possible that the mirrored system would exhibit lower turnover even with these (imperfectly) purified fuel strands. The difference should be particularly striking when comparing the results from using unpurified fuel strands, in which 5′ truncations are relatively common.

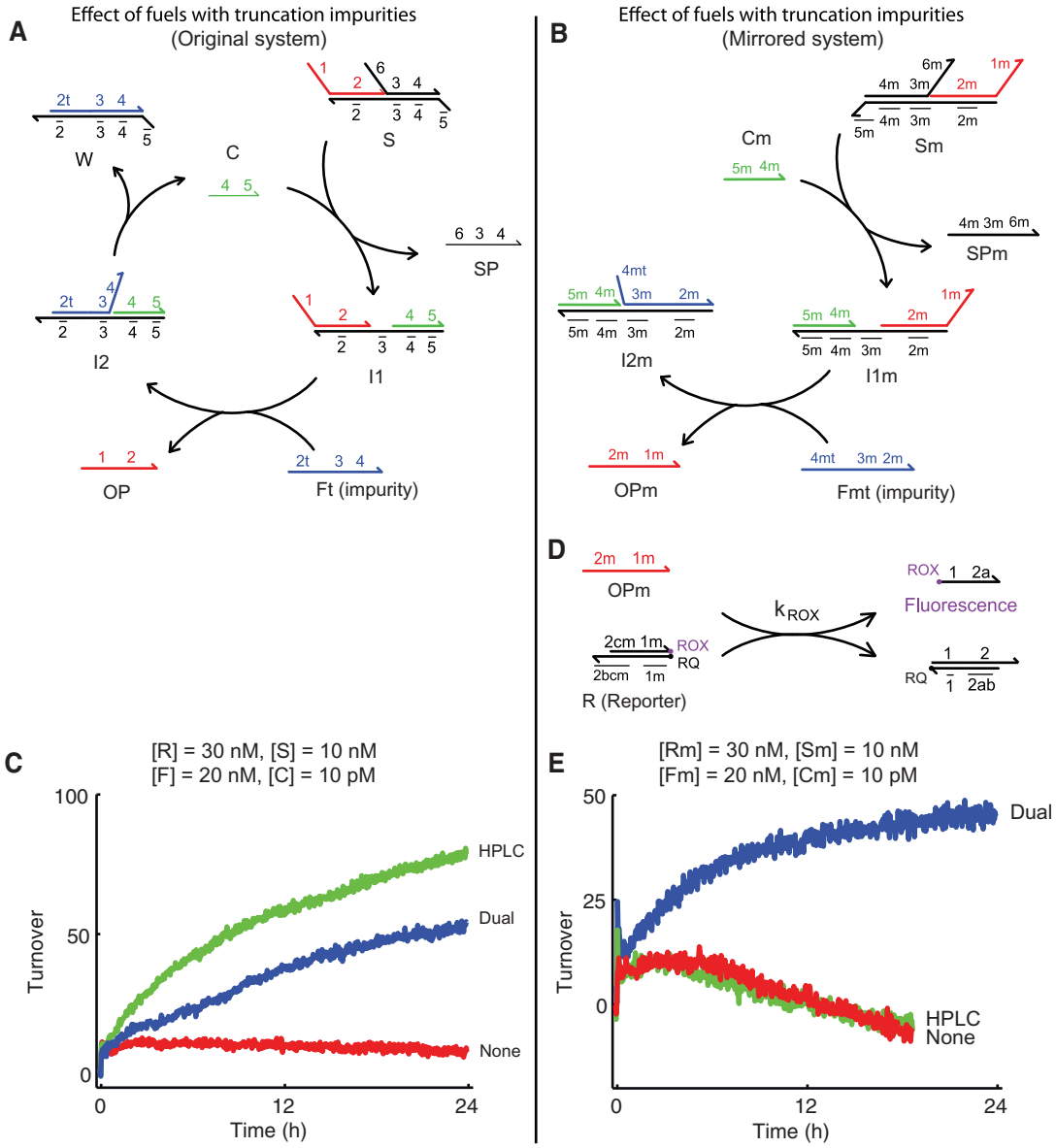

Figure 11. Effects of impurity and 5′/3′ orientation on maximum turnover. (**A**) Schematic of the original catalytic pathway when the fuel strand possesses an unintended 5′-Truncation. 5′-Truncated fuel strands *Ft* (due to synthesis errors and capping) do not hinder the designed catalytic pathway; this was shown in (9). (**B**) Schematic of the mirrored catalytic pathway when the fuel strand possesses an unintended 5′ truncation. 5′-Truncated fuel strands *Fmt* are unable to quickly displace catalyst *Cm* from intermediate *I2m*. Catalyst cannot be regenerated to catalyze other reactions, so turnover is severely limited. (**C**) Maximum turnover of the original catalyst design. Catalyst *C* is unpurified, and substrate complex *S* is PAGE purified from unpurified strands. Fuel molecules *F* are unpurified ('None' trace), HPLC purified ('HPLC'), or dual HPLC/PAGE-purified ('Dual'). HPLC- and dual HPLC/PAGE-purified fuels allow a maximum turnover of over 50, while unpurified fuel allows a maximum turnover of less than 10. Maximum turnover is calculated as in Figure 2B: the plotted turnover is the excess fluorescence signal of an experiment with stated concentration of *C* over that of an experiment lacking *C*, divided by (*C*). (**D**) The reporter complex for the mirror catalyst system; the strands for this reporter complex likewise mirror the sequence of the strands for the original catalyst system. (**E**) Maximum turnover measurement of the mirrored catalyst design. Catalyst *Cm* is unpurified, and substrate complex *Sm* is PAGE purified from unpurified strands. Fuel molecules *Fm* are unpurified ('None' trace), HPLC purified ('HPLC') or dual HPLC-PAGE purified ('Dual'). The maximum turnover of the mirrored catalyst system using dual HPLC/PAGE-purified fuel is seen to be roughly 45, while the HPLC purified and unpurified fuels allowed a maximum turnover of no more than 10. The spike near $t = 0$ is due to the subtractive nature of the method for calculating turnover, and is likely an artifact (this also exists in Figure 2B). Similarly, the decline in turnover at $t > 4$ h is likely also an artifact, due to the oligonucleotides in the uncatalyzed sample being at a slightly higher concentration (perhaps due to decreased adsorption of DNA to pipette tips) than that of the samples with catalyst.

We tested our hypothesis by performing experiments to measure the maximum turnover of the two systems, using different fuels. These experiments are similar to those shown in Figure 2, in that [*S*] >> [*C*] and [*Sm*] >> [*Cm*]. Unlike the previous experiments on maximum turnover, however, we used unpurified strands for all but the fuel strand, in order to isolate the effects of fuel purity on the maximum turnover. (The substrates *S* and *Sm* were PAGE purified from unpurified strands.)

The first surprise was that the maximum turnover of the original catalyst system was quite low when using unpurified fuel strands (less than 10, see Figure 11C).

Table 6. Domain sequences of mirrored sequences (Figure 11B)

Dom.	Sequence	Length (nt)
1m	5′-ACATCCTTTC-3′	10
2am	5′-GGCATTCAATCA-3′	12
2bm	5′-ACCTCT-3′	6
2cm	5′-GCATCC-3′	6
2m	5′-GGCATTCAATCAACCTCTGCATCC-3′	24
3m	5′-TCCC-3′	4
4m	5′-GCATCCCATAACTTAC-3′	16
5m	5′-ACCTCT-3′	6
6m	5′-TTATACTACATACACC-3′	16

This was unexpected because it was assumed that the difference between HPLC-purified strands and unpurified strands lay mostly in the absence of 5′-truncated strands in the former. Previous experimentation (2) demonstrated a simplified catalytic system in which *OP* was removed, along with the domain 2 of *F* and the domain $\bar{2}$ of *SL*. This suggests that catalysis can occur even with very large 5′ truncations in *F*, in which the entire domain 2 is missing (in this case, *SP* will be catalytically released, but not *OP*). One possible interpretation of the low turnover from using unpurified fuel is that the unpurified fuel strand *F* also possesses a significant fraction of strands with deletions near the 3′-end, which the HPLC purification was successfully able to remove.

In contrast, the exact same system using HPLC-purified and dual HPLC/PAGE-purified fuel strand exhibited maximum turnover of over 50 over 24 h (exact maximum turnover unknown). The results using the HPLC-purified fuel strand (green trace) quantitatively agree with the results in Figure 2B, implying that only high fuel strand purity is necessary for high turnover. It is unclear why the dual HPLC/PAGE-purified fuel strands yielded slower catalysis kinetics than HPLC-purified fuel strands.

As hypothesized, the maximum turnover of the mirrored system was significantly lower than that of the original system (Figure 11E and F). Using unpurified and HPLC-purified fuel strands yielded maximum turnover of only 5, while dual HPLC/PAGE-purified fuel strands allowed a maximum turnover of 45. The fact that the kinetics of the system using HPLC-purified *Fm* were nearly indistinguishable from that using unpurified *Fm* was surprising: it was expected that HPLC-purified *Fm* would contain only very minimal amounts of 5′ truncated *Fm*, and thus would exhibit maximum turnover closer to that of using the dual HPLC/PAGE-purified *Fm*.

Two general conclusions can be drawn from the results presented in this section: first, the purity of the fuel strand has a strong impact on the maximum turnover of the catalytic system, regardless of 5′/3′ orientation. Given that similar maximum turnover was achieved using purified and unpurified strands for the catalyst and substrate complex, this implies that high purity of these other strands (*C*, *SP*, *OP* and *SL*) are not as crucial for catalytic function and high turnover. Second, the 5′/3′ orientation of the original catalyst system allows higher maximum turnover, because of the asymmetric nature of oligonucleotide synthesis. However, it does appear that purification of strands and complexes may be able to alleviate this intrinsic disparity between orientations.

DISCUSSION

From the various experiments performed in this article, several observations made of the non-covalent DNA catalytic reaction are likely to be generalizable to all strand displacement-based DNA components.

First, whenever two single-stranded domains are located in close proximity, they are likely to interfere with each other's binding to their respective complements (Figure 4): catalytic activity was reduced when the catalyst possessed a 5′-overhang, presumably because of its interference with domain $\bar{3}$. Domain interference could be due to a combination of electrostatic repulsion, steric hindrance and non-specific binding; thus, interference is expected to be exacerbated when the lengths of the single-stranded domains increase. This observation is consistent with the slower kinetics of the previously reported allosteric catalyst (12) and hairpin-based catalytic structures (11).

Second, strand specificity can be controlled at the design level (Figures 6 and 7): the catalytic reaction is sensitive to single-base mutations in the catalyst, but robust to mutations in the fuel (except at the toehold domain 3). Sensitivity to differences in strand sequences is important for modularity of aqueous chemical systems, because of the large number of possible interactions that could occur in a well-mixed solution. Thus, DNA components displaying strong specificity will allow for larger and more complex synthetic chemical networks to be built. Previously, a DNA circuit with 11 components and 6 layers deep was experimentally demonstrated (4); in theory, DNA circuits can be scaled up further and integrate more different components (26).

Third, strands do not need to be high purity in order to function (Figure 8): the catalytic reaction proceeded qualitatively and quantitatively similarly with strands of <80% purity as with strands of >90% purity. Impurity tolerance greatly reduces the time and cost of DNA component preparation and synthesis.

Fourth, strand displacement-based DNA components are robust to certain types of molecular noise: the DNA catalytic reaction characterized here, though sensitive to excess poly-N ssDNA, was robust to large amounts of poly-H ssDNA in solution. Just as with certain electronic modules that are robust to pink or gray but not white noise, this feature of strand displacement-based DNA systems should be considered in the design process to mitigate the effects of noise. For purely synthetic networks in which all nucleic acid molecules are designed, sequences can be chosen to take advantage of the selective noise robustness of strand displacement-based systems.

Fifth, the use of all four nucleotides in the sequence design of domains is seen to significantly slow down the kinetics of DNA toehold-mediated strand displacement, despite the involved strands not possessing any significant predicted secondary structure. This could possibly be due

to fleeting intra- and inter-molecular interactions not captured by thermodynamics, but may be revealed through simulation of DNA strand displacement kinetics at elementary step resolution (28).

Finally, the fact that DNA oligonucleotide synthesis proceeds from the 3′- to the 5′-end leads to an intrinsic asymmetry in the fidelity of the bases for synthetic DNA, with the 5′ bases being more likely to be truncated or deleted. This, in turn, leads to a 'preferred' 5′/3′ orientation when designing catalytic reactions, in which lower purity strands can be used to achieve high catalytic turnover.

With the results presented herein, we remain cautiously optimistic about the future of strand displacement-based DNA circuits and networks. While integration of these constructions *in vivo* will likely require further technological breakthroughs, strand displacement-based DNA components may immediately be used to facilitate control of synthetic chemistry.

ACKNOWLEDGEMENTS

We thank Georg Seelig and David Soloveichik for insightful discussions.

FUNDING

National Science Foundation (grants 0506468, 0622254, 0533064 and 0728703 to D.Y.Z. and E.W.) Fannie and John Hertz Foundation (to D.Y.Z.). Funding for open access charge: the National Science Foundation grant 0728703.

Conflict of interest statement. None declared.

REFERENCES

1. SantaLucia,J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
2. Zhang,D.Y. and Winfree,E. (2009) Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.*, **131**, 17303–17314.
3. Yurke,B. and Mills,A.P. (2003) Using DNA to power nanostructures. *Genet. Prog. Evol. Mach.*, **4**, 111–122.
4. Seelig,G., Soloveichik,D., Zhang,D.Y. and Winfree,E. (2006) Enzyme-free nucleic acid logic circuits. *Science*, **314**, 1585–1588.
5. Hagiya,M., Yaegashi,S. and Takahashi,K. (2006) Computing with hairpins and secondary structures of DNA. In Rozenberg,G., Back,Th., Eiben,A.E., Kok,J.N. and Spaink,H.P. (eds), *Nanotechnology: Science and Computation*. Springer, Berlin, Heidelberg, pp. 293–308.
6. Frezza,B.M., Cockroft,S.L. and Ghadiri,M.R. (2007) Modular multi-level circuits from immobilized DNA-based logic gates. *J. Am. Chem. Soc.*, **129**, 14875–14879.
7. Yurke,B., Turberfield,A.J., Mills,A.P., Simmel,F.C. and Neumann,J.L. (2000) A DNA-fuelled molecular machine made of DNA. *Nature*, **406**, 605–608.
8. Lubrich,D., Lin,J. and Yan,J. (2008) A contractile DNA machine. *Angew. Chem. Int. Ed.*, **47**, 7026–7028.
9. Zhang,D.Y., Turberfield,A.J., Yurke,B. and Winfree,E. (2007) Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, **318**, 1121–1125.
10. Dirks,R.M. and Pierce,N.A. (2004) Triggered amplification by hybridization chain reaction. *Proc. Natl Acad. Sci. USA*, **101**, 15275–15278.
11. Yin,P., Choi,H.M.T., Calvert,C.R. and Pierce,N.A. (2008) Programming biomolecular self-assembly pathways. *Nature*, **451**, 318–322.
12. Zhang,D.Y. and Winfree,E. (2008) Dynamic allosteric control of noncovalent DNA catalysis reactions. *J. Am. Chem. Soc.*, **130**, 13921–13926.
13. Seelig,G., Yurke,B. and Winfree,E. (2006) Catalyzed relaxation of a metastable DNA fuel. *J. Am. Chem. Soc.*, **128**, 12211–12220.
14. Bois,J.S., Venkataraman,S., Choi,H.M.T., Spakowitz,A.J., Wang,Z.G. and Pierce,N.A. (2006) Topological constraints in nucleic acid hybridization kinetics. *Nucleic Acids Res.*, **33**, 4090–4095.
15. Turberfield,A.J., Mitchell,J.C., Yurke,B., Mills,A.P., Blakey,M.I. and Simmel,F.C. (2003) DNA fuel for free-running nanomachines. *Phys. Rev. Lett.*, **90**, 118102.
16. Gartner,Z.J., Tse,B.N., Grubina,R., Doyon,J.B., Snyder,T.M. and Liu,D.R. (2004) DNA-templated organic synthesis and selection of a library of macrocycles. *Science*, **305**, 1601–1605.
17. Aldaye,F.A., Palmer,A.L. and Sleiman,H.F. (2008) Assembling materials with DNA as the guide. *Science*, **321**, 1795–1799.
18. Clo,E., Snyder,J.W., Voigt,N.V., Ogilby,P.R. and Gothelf,K.V. (2006) DNA-programmed control of photosensitized singlet oxygen production. *J. Am. Chem. Soc.*, **128**, 4200–4201.
19. Feldkamp,U., Sacca,B. and Niemeyer,C.M. (2009) Dendritic DNA building blocks for amplified detection assays and biomaterials. *Angew. Chem. Int. Ed.*, **48**, 5996–6000.
20. Marras,S.A.E., Kramer,F.R. and Tyagi,S. (2002) Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Res.*, **30**, e122.
21. Gao,Y., Wolf,L.K. and Georgiadis,R.M. (2006) Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. *Nucleic Acids Res.*, **34**, 3370–3377.
22. Puglisi,J.D. and Tinoco,I. (1989) Absorbance melting curves of RNA. *Methods Enzymol.*, **180**, 304–325.
23. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
24. Bloomfield,V.A., Crothers,D.M. and Tinoco,I. (2000) *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, Sausalito, CA.
25. Temsamani,J., Kubert,M. and Arawal,S. (1995) Sequence identity of the n-1 product of a synthetic oligonucleotide. *Nucleic Acids Res.*, **23**, 1841–1844.
26. Qian,L. and Winfree,E. (2009) A simple DNA gate motif for synthesizing large-scale circuits. In Goel,A., Simmel,F.C. and Sosik,P. (eds), *DNA Computing: 14th International Meeting on DNA Computing*. Springer, Berlin, Heidelberg, pp. 70–89.
27. Dirks,R.M., Bois,J.S., Schaeffer,J.M., Winfree,E. and Pierce,N.A. (2007) Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev.*, **49**, 65–88.
28. Flamm,C., Fontana,W., Hofacker,I.L. and Schuster,P. (2000) RNA folding at elementary step resolution. *RNA*, **6**, 325–338.